Theorems and Models in Political Theory: an Application to Pettit on Popular Control

Sean Ingham*

February 18, 2015†

1.

In an early passage from *A Theory of Justice*, Rawls suggests that not only will moral and political philosophy "involve principles and theoretical constructions which go much beyond the norms and standards cited in everyday life," but they may "eventually require fairly sophisticated mathematics as well."[1] This second claim would find assent among only a small minority of political theorists and philosophers today (or then), if the near total absence of mathematics from the discipline is any guide. But why should this be so? Many political theorists, in

*Assistant Professor of Political Science, University of Georgia. Email: ingham@uga.edu.

†I am grateful to Lucas Stanczyk and two anonymous reviewers for carefully reading and suggesting improvements to earlier versions of the manuscript. Any remaining mistakes are my own.

[1]Rawls (1971, 47). Interestingly, these remarks were preserved in the 1999 revised edition, even though Rawls chose to remove the next sentence: "This is to be expected, since on the contract view the theory of justice is part of the theory of rational choice" (Rawls, 1999, 41, 42).

particular those who see themselves as occupying the more "analytical" wing of the discipline, consider the careful definition of terms and the construction of deductive arguments as an important part of their craft. It would be surprising if mathematics never had a useful role to play in political theory, given this understanding of the enterprise.

This essay presents an example of how political theory can benefit from mathematics, *i.e.*, formal definitions and deductive reasoning. I begin in the next section with observations on Phillip Pettit's recent account of popular control. These observations are the stuff of "conventional" political theory—I claim he conflates two distinct conditions in his conception of popular control—and they neither require nor benefit (much) from formal constructions. But they raise questions that would be all but impossible to answer without formal definitions and arguments. In section 3, I present formal definitions that facilitate this inquiry. In section 4, the inquiry proceeds through increasingly denser thickets of logical complexity, culminating in a theorem that is similar in its underlying logical structure to theorems on preference-aggregation, like the famous theorems of Arrow or Gibbard and Satterthwaite: popular control, as Pettit models it, implies that either the government is sometimes perversely responsive to citizens' normative attitudes or there is a single citizen whose normative attitudes unilaterally fix—in the manner of a "dictator"—government policy.

The argument requires only elementary mathematical concepts, like that of a set and a function, so it is not a display of "sophisticated" mathematics per se. But with this qualification, I hope it will illustrate the plausibility of Rawls's

suggestion.

2.

Pettit (2013) offers a novel republican theory and model of democracy. The theory consists of an account of popular control over government, which Pettit argues is necessary for non-domination and legitimacy. The model is meant to show how a certain kind of deliberative democracy could establish popular control. But the theory and model seem to operate with distinct interpretations of popular control. The theory sets up an ideal of popular control that is more stringent and demanding than the form of popular control that the model describes as feasible, or so I will argue.

Pettit's general notion of control takes it to be influence that imposes a direction or pattern on a process. For example, when I cause chaos by standing in the middle of an intersection waving my hands, I influence but do not control the flow of traffic; when the police officer does the same, his influence imposes a pattern on the flow of traffic and can therefore be described as control (Pettit, 2013, 154). Popular control must involve both "popular influence and popular direction" (Pettit, 2013, 168). When Pettit describes the kind of popular control required for non-domination and legitimacy, he argues that "people should share equally both in exercising influence over government and in determining what direction that influence is to impose" (Pettit, 2013, 169). This latter requirement means that "the direction their influence imposes [...] is required to be one that

each is ready to accept." Pettit adds an important clarification in a footnote to the claim:

> Saying that something is acceptable often has normative significance [...], implying that it is such that people ought to accept it. Here and throughout this book, the word has a non-normative sense, implying that the object or policy or whatever is such that people are disposed to accept it; they find it acceptable, as we say (Pettit, 2013, 170).

Pettit discusses other conditions that a system of popular control must satisfy, if it is to eliminate or minimize citizens' domination by the state, but my focus will be on his idea that it must involve a system of influence that "[gives] the state an equally acceptable direction—that is, a direction that [citizens] are all actually disposed to accept" (Pettit, 2013, 170).

One might doubt how this condition could ever be met. For any "direction" to policy-making, surely there will always be at least one citizen who is not disposed to accept it. Pettit acknowledges that the requirement of universal acceptability "may not be satisfiable amongst fanatics or zealots who insist on special treatment: say, the privileging of their religion or ethnicity." But it can be satisfied, he holds, among citizens "who are willing to live on equal terms with others" (Pettit, 2013, 170). Restricting in this manner the set of people to whom the direction of government must be acceptable goes some way towards meeting the worry. But only some. It is not obvious why some mutually acceptable direction of policy should be assumed to exist in a diverse society divided by

class and other cleavages, even if we assume that everyone wishes to live on equal terms with each other. Why should we expect there to be a direction that policy-making could take which *every* citizen—the Christian fundamentalist as well as the secular liberal, the hedge fund manager as well as the assembly-line worker—will be disposed to accept, even granting that they are all prepared to live on equal terms with each other? What sharpens the point of this question is the non-normative meaning of "acceptable." As Pettit stresses in the footnote quoted above, what matters is not whether there is a direction that everyone who is willing to live on equal terms with others *should* be ready to accept. Rather, what matters is whether there is a direction that everyone who is willing to live on equal terms with others *is* ready to accept.

That "acceptability" has a non-normative meaning in Pettit's interpretation of popular control is in itself a virtue. Control over something should make it sensitive to one's *actual* attitudes, not the attitudes that one ought to have. But the worry is that these truth-conditions of "acceptable to citizen k" combine with the universal quantifier ("*every* citizen k") to make Pettit's formulation of popular control an impossible ideal.

Indeed, when Pettit turns to the model of a deliberative democracy that is meant to illustrate the feasibility of popular control, he shifts towards a less demanding requirement. The model shows how a system of popular influence might succeed at imposing a direction on policy-making that is consistent with universally accepted policy-making norms. As I explain below, this weaker constraint may be satisfied even if not everyone finds the direction of policy accept-

able. In fact, it can be satisfied even if everyone finds the direction of policy unacceptable.

In Pettit's model of democracy, the mechanism underlying popular control is deliberation governed by "a norm to the effect that participants should only offer considerations for or against a policy that all can regard as relevant."

> The idea is that any considerations adduced should help to make the policy acceptable to everyone, given shared assumptions about the dispositions of each... The considerations adduced should count as relevant according to everyone's views but according to everyone's views as they actually are or can be brought to be, not according to everyone's views as in some sense they ought ideally to be.
>
> ... The only considerations they can invoke are ones that argue on all sides – though perhaps with a varying force or weight – for accepting the policy supported; they must count with everyone as grounds that it is at least relevant or pertinent to adduce in arguing for or against a policy (Pettit, 2013, 253, 254).

Where this deliberative norm is in place, "it will force people to interact in a manner that gives rise to shared policy-making norms." Even where there are partisan divides over policy, "the requirement for the parties on each side to defend their partisan proposals in multi-partisan terms ensures that as they construct their different programmes, they will lay down a foundation of common ground between them." The partisan dissensus "will secure or reinforce the

norms of argument that the disagreement drives the different sides to identify" (Pettit, 2013, 261).

Gradually, generally accepted policy-making norms should emerge as a result of this kind of deliberation, and under the right conditions—the right democratic institutions and culture—these norms should constrain policy. Their constraining effect should ensure that two conditions are fulfilled.

> First, a policy-condition: no policies are left on the table as possible candidates for adoption in any domain, if they are in violation of the norms. And second, a process-condition: no process for selecting between candidates that survive that first cut – and no process for selecting between rival processes – is employed, if its employment in that domain would be in violation of the norms (Pettit, 2013, 265).

Suppose that everything goes according to script, and the government is prevented from choosing any policy that violates any of these generally accepted public norms of policy-making. The system of popular influence imposes a "norm-complying" direction on policy-making, a direction that complies with all universally accepted policy-making norms. Can we conclude that it has imposed a direction that everyone is disposed to accept? We cannot. Some citizens may not be disposed to accept it even if they agree that it does not violate any of the publicly recognized, universally accepted policy-making norms.

Some citizens may find a policy-direction unacceptable, even though it respects all publicly recognized, universally accepted policy-making norms, because it violates principles that they—but not all citizens—accept as binding con-

straints on government policy. Someone who accepts Rawls's difference principle may not be disposed to accept the direction that policy has taken in the United States over the last forty years. Nonetheless, the direction policy has taken may be compatible with all those policy-making norms that *everyone* accepts. For not everyone accepts the difference principle.

In fact, a system of popular influence may impose a direction on policy that complies with all universally accepted policy-making norms, even though *no one* is disposed to accept it. For each person, it may violate a policy-making norm that she, but not everyone, accepts; she may not regard it as acceptable for this reason; yet there may be no norm which everyone accepts and which condemns the direction policy has taken.

The model therefore delivers a form of popular control that is different from what the theory promises. The theory promises a system of popular influence that imposes a direction on policy-making that everyone is disposed to accept; the model delivers a system of popular influence that imposes a direction on policy-making that is compatible with norms that everyone accepts. The constraint on government that Pettit's theory formulates is not a constraint that the government in his model of a deliberative democracy need satisfy. Government may take a direction that satisfies all universally accepted policy-making norms, but which not everyone is disposed to accept.

One thought that seems to facilitate Pettit's shift to the weaker requirement is the thought that all citizens will want government to be forced to comply with universally accepted policy-making norms. If a system of popular influ-

ence has this effect, the effect will be welcome—indeed, equally acceptable to all. As Pettit claims in the concluding pages of his discussion of the model, "the effect of a system of popular influence in forcing government to comply with accepted policy-making norms is bound to be equally acceptable to all participants" (Pettit, 2013, 280). But this is not the same thing as their influence imposing a direction on government that is acceptable to each. For example, a student may accept the constraining effect of university policy on a professor's determination of his grade , but not accept the professor's determination of his grade. Accepting the constraining effect of universally accepted norms on the direction policy takes is not the same as accepting the direction policy takes. Yet, according to the theory of democracy, the system of popular influence "has to give the state an equally acceptable direction—that is, a direction that [citizens] are all actually disposed to accept." It is not enough for it to have a constraining effect that all citizens welcome (Pettit, 2013, 170).

So there is the logical gap between the policy-directions compatible with universally accepted policy-making norms and the policy-directions that are acceptable to each. The next section explores the implications of this observation with the help of a formalization of Pettit's model. The first payoff of the formal model is the insight that changes in citizens' normative attitudes may have strange, perverse effects on the behavior of a government subject to popular control.

3.

In this section I present a formal model that consists in set-theoretic definitions of policy-making norms, the norms a citizen accepts, government's subjection to these norms, and so on—in short, all of the concepts under discussion in the previous section. The definitions do not invoke any mathematical concepts other than those of a set, a collection of sets, the intersection of sets, etc. Why go through the exercise of formalizing Pettit's model of popular control, if the model consists in nothing more than set-theoretic definitions of the same concepts that figure in the non-technical version of the model?

These definitions and the symbols for the sets that figure in them serve a purpose similar to that of symbols for arithmetical operations ($+$, $-$, etc.). The proposition that the sum of two numbers, when multiplied by itself, equals the first number multiplied by itself plus the second number multiplied by itself plus twice the product of the two numbers is hard to understand, much less evaluate, when it is expressed in "plain English." It is easier to understand when expressed so:

$$(x + y)^2 = x^2 + 2xy + y^2.$$

Just as special symbols are aids in mental computations of products and sums, the formal representation of norms and policies as sets, and the introduction of special symbols for these sets, are mental aids that will help us keep track of the moving parts of questions about popular control. Obviously, not every question about popular control requires such aids, just as not every arithmetical

computation requires pen and paper. But, as I hope this section will demonstrate, there are some questions that hold substantive interest and also exhibit enough complexity to justify the introduction of the formal model.

These questions are variations on the following: in what sense does popular control, in the sense of subjection to accepted policy-making norms, imply that government is *responsive* to citizens' normative attitudes? The first benefit of the formal model is that it allows us to demonstrate that popular control in this sense is—in principle—compatible with the government being "responsive" in a perverse way: as a result of a change in citizens' attitudes that involves nothing more than some citizens coming to find the government's current policy acceptable, the government may be led to adopt a different policy that these citizens consider unacceptable.

3.1

We will let X refer to the set of all feasible options open to the government. (These are the various "directions" which popular influence may impose on government.) Let $N = \{1, 2, \ldots, n\}$ designate the set of citizens.

A given policy-making norm permits the government to choose some subset of the options in X. We will therefore represent a policy-making norm by the subset $P \subseteq X$ of the feasible options that are compatible with it. To take a simple example, suppose the government faces a choice between three options affecting the basic structure of society, $X = \{x, y, z\}$, and Rawls's principle of fair equality of opportunity is compatible with x and y, but not z. Then,

within the formal model, the fair equality of opportunity principle is represented by the subset $\{x, y\}$. The set of policy-making norms that a citizen accepts is then represented as a set of subsets. For example, if Rawls's difference principle permits only x, then the normative attitudes of the person who accepts just the difference principle and the fair equality principle are represented by the set $\{\{x\}, \{x, y\}\}$.

This abstract representation of the norms that a citizen accepts leaves out a lot of information, but that is the point. The goal is to represent only the information that is relevant to the following arguments and to leave out any information that would only clutter the exposition and distract us from what is driving the conclusions.

For each $i \in N$, we will let $\mathcal{P}_i$ denote the collection of sets that represents the norms that citizen i accepts. To continue with our simple example, suppose that $N = \{1, 2\}$ and citizen 1 accepts the difference principle and fair equality of opportunity principle, but citizen 2 accepts only the fair equality of opportunity principle. Then we would have:

$$\mathcal{P}_1 = \{\{x\}, \{x, y\}\},$$
$$\mathcal{P}_2 = \{\{x, y\}\}.$$

That $\{x\} \in \mathcal{P}_1$ means that citizen 1 accepts a norm (the difference principle) which permits only x—equivalently, it rules out any option z such that $z \neq x$. That, additionally, $\{x, y\} \in \mathcal{P}_1$ means that citizen 1 also accepts a norm (fair

equality of opportunity) that permits only x or y—equivalently, it rules out any option z such that $z \notin \{x, y\}$.

If an option x is permitted by every policy-making norm that she accepts, then I will say that she finds it acceptable; if it violates a policy-making norm she accepts, I will say that she finds it unacceptable. Let us assume that each citizen can identify at least one acceptable option (which is plausible when X comprises *every* feasible option, as it does here). Formally, the assumption is that for each citizen, the intersection of the norms she accepts is nonempty: $\cap_{P \in \mathcal{P}_i} P \neq \emptyset$ for each citizen $i \in N$. Each person's normative commitments are jointly compatible with at least one of the feasible options.

The simple example just given illustrates the observation from the previous section: y is compatible with all universally accepted policy-making norms, because the only universally accepted norm is $\{x, y\}$, the fair equality of opportunity principle. But citizen 1 is not disposed to accept it because it violates the difference principle, which she accepts.

So far we have simply used this formal set-theoretic language to articulate a thought that some might consider to have been already satisfactorily expressed in plain English. Its restatement alone might not justify the effort of formalizing Pettit's model of popular control. Let me now try to make the case that this exercise yields additional benefits that do justify it.

I claimed above that the consequence of some citizens coming to find an option acceptable may be that the government chooses a different option that they find unacceptable, even if (and partly because) the government is subject to

popular control as Pettit models it. With the formal model in place, this claim is easy to verify.

As before, let $X = \{x, y, z\}$ and $N = \{1, 2\}$. Suppose all three options are compatible with the principle of efficiency, which is therefore represented by the set X. Fair equality of opportunity and the difference principle correspond, as before, to $\{x, y\}$ and $\{x\}$, respectively. And let us suppose that there is an additional norm of meritocracy, $\{y\}$, which rules out the kind of redistribution involved in x and permits only y. Suppose that initially the citizens' normative attitudes are given by

$$\mathcal{P}_1 = \{\{x, y\}, X\},$$
$$\mathcal{P}_2 = \{\{y\}, \{x, y\}, X\},$$

That is, they both accept the principle of efficiency and the principle of fair equality of opportunity, and citizen 2 also accepts the meritocratic norm. Thus, the government, being constrained to choose an option compatible with each universally accepted policy-making norm, is constrained to choose either x or y.

Let us suppose the government has its own preferences over the three options, which determine how it chooses when the constraints of popular control leave it with some discretion: it prefers z to x and prefers both of these to y. Given its preferences, it chooses x. Note that citizen 2 considers this option unacceptable, because it violates the norm of meritocracy.

Now consider a second scenario in which the second citizen—perhaps after reading Arneson (1999)—has come to reject the meritocratic norm as well as the fair equality of opportunity principle, but has come to affirm the difference principle.

$$\mathcal{P}_1' = \{\{x, y\}, X\},$$
$$\mathcal{P}_2' = \{\{x\}, X\}.$$

Note that in light of the changes in citizen 2's normative attitudes, she now accepts the government's initial choice of x—indeed, she now considers it uniquely acceptable—whereas before she considered it unacceptable. But now the only generally accepted policy-making norm is X, the principle of efficiency. Thus the government is free to choose z, which it prefers to the other two options. So the effect of citizen 2 revising her normative attitudes and coming to regard x as acceptable—where she previously considered it unacceptable—is that the government is permitted to choose and does choose z—which she previously regarded and still regards as unacceptable—instead of x.

That Pettit's system of norm-based regulation could operate in this manner raises a question about the plausibility of describing it as a system of popular control. Not only does norm-based regulation not compel the government to choose options that everyone finds acceptable. But it also allows government policy to be perversely responsive to changes in citizens' normative attitudes in the following sense. The attitudes of the second citizen change and this change

has an effect on the government's choice, but the effect is unwelcome from citizen 2's point of view: the outcome would be better, from her point of view, if the change in her attitudes were to go unregistered. For the new option fares worse than the old, relative to the normative standards that she now accepts. That seems like a perverse feature of the relationship between citizens' normative attitudes and the object of their putative control. Intuitively, popular control implies responsiveness to citizens' attitudes, but the object of their control should never be "negatively" responsive to their attitudes, as it is here.

Now, perhaps it is only in virtue of the government's preferences, but not its subjection to norm-based popular control per se, that causes the government's chosen option to be perversely responsive in this manner to the change in citizen 2's normative attitudes. Perhaps subjection to norm-based popular control would not give rise to this phenomenon if the government had different preferences or faced additional constraints on its choices. After all, in the example the government *could* continue to choose x without violating any generally accepted norm. It refrains from doing so only because of its preference for z, which is also compatible with accepted norms in the second scenario. In doing so, it may seem to be violating a second-order policy-making norm concerning how it should respond to shifts in citizens' normative attitudes. How might one formulate the second-order norm that it violates here?

Here is one thought. Let us say that a citizen recognizes a reason for choosing x instead of y if she accepts a norm that permits x but not y. A second-order norm that might capture the intuition about positive responsiveness is this:

> **(Positive responsiveness)** Government policy (or its direction) should not shift from x to y, unless some citizen has either come to recognize a reason for choosing y instead of x or ceased to recognize a reason for choosing x instead of y.

This condition is equivalent to saying that if citizens' normative attitudes change, but no one's attitudes become more favorable towards y relative to x—that is, no one has come to recognize a new reason for choosing y instead of x, and no one has ceased to recognize a reason for choosing x instead of y—then the effect of the changes in their normative attitudes cannot be that the government changes its choice from x to y.[2]

In the previous example, this second-order norm would prevent the choice of z in the second scenario if the government chooses x in the first scenario. For although the citizens' normative attitudes change, there is no citizen who comes to recognize a reason for choosing z instead of x and no citizen who ceases to recognize a reason for choosing x instead of z. The first citizen's attitudes do not change. The second citizen still recognizes every reason for choosing x instead of z that she recognized before—namely the fair equality of opportunity principle, $\{x, y\}$. And she has not come to recognize any reason for choosing z instead of x—there is no new norm P such that she accepts P in the second scenario ($P \in \mathcal{P}_2'$) even though she did not accept it in the first ($P \notin \mathcal{P}_2$) and that rules out x but not z ($z \in P$ but $x \notin P$). Positive responsiveness therefore

[2] In other words, positive responsiveness means that the condition following *unless* in the definition is a necessary—but not necessarily sufficient—condition for government policy shifting from x to y.

requires that the government not choose z, given that it chose x in the first scenario. (It also requires that it not choose y.)

In this example, it is possible for the government to respect this second-order norm at the same time that it respects every generally accepted first-order policy-making norm; it satisfies both constraints if it chooses x in both scenarios. But how can we know whether it is possible *in general* to satisfy both constraints? That is, how can we be sure that the mutual satisfiability of these two constraints is not due to special features of the examples? An extension of the formal model will help us answer this question.

3.2

In this section I add to the existing model a representation of how citizens' normative attitudes influence the direction of government, that is, its choice from the (finite) set of options, X. This addition will allow us to determine whether it is possible *in general* for a government to be both subject to generally accepted public norms and positively responsive to changes in citizens' attitudes. The conclusion will be that if X contains at least three options, then a government subject to norm-based popular control is positively responsive to changes in citizens' normative attitudes only if there is some citizen—a "dictator"—whose attitudes can unilaterally determine government's behavior.

Since positive responsiveness concerns the manner in which citizens' normative attitudes influence government policy (or the direction of government), it will be useful to incorporate a representation of this relationship into the

model. Each citizen's normative attitudes are represented by a (non-empty) collection of subsets of X such that the subsets have a non-empty intersection.[3] Let $\mathbf{P}$ denote the set of all such collections.[4] A full specification of all citizens' normative attitudes is therefore an n-tuple $\mathcal{P} = (\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_n) \in \mathbf{P}^n$. The causal relationship between the government's choice and citizens' normative attitudes—the manner in which the latter influence the former—can be represented as a function $g : \mathbf{P}^n \to X$, where $g(\mathcal{P}) \in X$ is the option chosen when citizens' normative attitudes are $\mathcal{P} = (\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_n)$. Positive responsiveness and norm-based popular control can now be expressed as conditions that this function may or may not satisfy.

Definition 1 (positive responsiveness)**.** *A causal relationship $g : \mathbf{P} \to X$ satisfies* positive responsiveness *if, for all $\mathcal{P}, \mathcal{P}' \in \mathbf{P}^n$ and $x, y \in X$, if $g(\mathcal{P}) = x$ and $g(\mathcal{P}') = y$, then there is some $i \in N$ such that either*

(i) there is a $Q \in \mathcal{P}_i'$ with $y \in Q$ and $x \notin Q$, but for all $P \in \mathcal{P}_i$, $y \in P$ only if $x \in P$; or

(ii) there is a $Q \in \mathcal{P}_i$ with $x \in Q$ and $y \notin Q$, but for all $P \in \mathcal{P}_i'$, $x \in P$ only if $y \in P$.

This is just a restatement of the previous definition of positive responsiveness. In words, government policy only changes from x to y if there is some

[3] Recall the assumption from the previous section that each citizen can always identify one option as acceptable.

[4] For example, if $X = \{a, b\}$, then $\mathbf{P}$ contains five elements: $\{\{a\}\}$, $\{\{b\}\}$, $\{X\}$, $\{\{a\}, X\}$, and $\{\{b\}, X\}$.

citizen who either (i) has come to accept a norm that permits y but not x (where she previously accepted no such norm); or (ii) no longer accepts a norm that permits x but not y (where she previously did accept some such norm). That is, the government's choice changes from x to y only if some citizen has either (i) come to recognize a reason for choosing y instead of x or (ii) ceased to recognize a reason for choosing x instead of y.

Definition 2 (norm-based popular control)**.** *A causal relationship* $g : \mathbf{P}^n \to X$ *satisfies* norm-based popular control *if, for all* $\mathcal{P} \in \mathbf{P}^n$*, we have* $g(\mathcal{P}) \in P$ *whenever* $P \in \mathcal{P}_i$ *for all* $i \in N$*.*

We wish to know whether there is any function $g : \mathbf{P}^n \to X$ that satisfies both of these conditions. The answer is that, if there are at least three options in X, then g satisfies norm-based popular control and positive responsiveness only if it is "dictatorial" in the sense of the following definition.

Definition 3 (dictatorship)**.** *A causal relationship* $g : \mathbf{P}^n \to X$ *is* dictatorial *if there is some* $k \in N$ *such that* $g(\mathcal{P}) \in \cap_{P \in \mathcal{P}_k} P$ *for all* $\mathcal{P} \in \mathbf{P}^n$*.*

Suppose there is a "dictator": a citizen such that, for any specification $\mathcal{P}$ of citizens' normative attitudes, government policy must be permitted by all of the norms that she accepts—it must be contained in the intersection of the norms she accepts. Then, even if everyone else believes some policy x is permissible—indeed, even if everyone else believes x is uniquely acceptable—the government will not choose x if x violates one of the norms that the dictator accepts. And even if everyone else believes some policy x is unacceptable, and they can all

agree on an acceptable alternative, the government must choose x if the dictator believes that x is uniquely acceptable. Thus, the power that some single citizen will have when the relationship g is dictatorial is considerable indeed. As it turns out, some citizen must have this power if government is positively responsive to citizens' normative attitudes and subject to popular control.

Theorem 1. *Assume* $|X| \geq 3$. *If* $g : \mathbf{P}^n \to X$ *satisfies positive responsiveness and norm-based popular control, then it is dictatorial.*

The interested reader can find a proof of the theorem in the appendix.

No more than one person can be a dictator, so a corollary of the theorem is that citizens' normative attitudes have unequal influence over government if the relationship between these attitudes and government policy satisfies popular control and positive responsiveness.

Like Arrow's impossibility theorem and the related Gibbard-Satterthwaite theorem, the theorem states that a certain kind of aggregation function is "dictatorial" if it satisfies other seemingly attractive properties (Arrow, 1951; Gibbard, 1973; Satterthwaite, 1975). In fact, the connection between the theorem and these results is deeper: as shown in the appendix, the theorem is a logical corollary of the Gibbard-Satterthwaite theorem (which can itself be derived from Arrow's theorem). As it is conventionally interpreted, the Gibbard-Satterthwaite theorem answers a question about voting rules—are there non-manipulable voting rules? On its face, it has nothing to do with the question under examination here—can government be both positively responsive to citizens' attitudes towards policy-making norms and subject to norm-based popular control, as

Pettit models it? Yet it turns out that the two inquiries are logically linked; from the answer to the first question one can derive the answer to the second. This finding illustrates a frequent benefit of formal models: they reveal logical parallels between seemingly unrelated inquiries. It is also another piece of evidence for an expansive view of the significance of Arrow's theorem and the Gibbard-Satterthwaite theorem: these results are not specifically about *preference* aggregation—notwithstanding the story that usually accompanies them—but rather a more general kind of aggregation problem (Patty and Penn, 2014). The aggregation of individuals' dispositions to accept norms is, it appears, one more example of this more general problem.

What conclusions should we draw from the theorem? We should not reject Pettit's interpretation of popular control on the basis of the theorem. For the positive responsiveness requirement is surely less compelling than either norm-based popular control or equality. If any of these requirements is shown by the theorem to be unreasonable, it is the requirement of positive responsiveness.

This requirement was formulated for a purpose, so its rejection comes at a cost. It was meant to capture the intuitive judgment that a government subject to popular control should not respond "negatively" to citizens' attitudes. In the example from section 3.1, a citizen's coming to find the government's current policy acceptable should not cause the government to change its policy to one that she considers unacceptable. Positive responsiveness does indeed capture this intuitive judgement, but it combines with the assumption of norm-based popular control to yield inegalitarian conclusions. The next question to ask

is whether there are alternatives to the positive responsiveness property that would capture our intuitions in this example, or whether the intuitive judgments are simply misguided. The formal model and theorem should be the point of departure for future inquiry into such questions.

One might also respond to the theorem by rejecting the assumption—implicit in this framework as well as Pettit's model—that popular control over government consists in a relationship between government and the policy-making norms that citizens' accept. Perhaps it consists instead in a relationship between the object of popular control and individual-level data that contain more information than merely which policy-making norms citizens accept. To take just one possibility, one might argue that popular control consists in a relationship between policy, on the one hand, and citizens' policy preferences and the strength with which they hold their preferences, on the other (Ingham, Forthcoming). As is well-known, one can sometimes escape Arrow-like impossibility results by enriching the informational environment.[5]

5.

There are arguably also conclusions to draw, not from the theorem itself but from the fact that one can get such a theorem once one represents government's subjection to accepted norms and its responsiveness to citizens' normative judgments in terms of sets and functions. The conclusion we should entertain is that Rawls was right to suggest that mathematics might sometimes have a role to

[5]For an accessible discussion and references to other literature on the topic, see Sen (2014).

play in political theory. Is there any plausible understanding of political theory that justifies skepticism towards Rawls's conjecture? Is there any plausible basis for drawing the disciplinary boundaries of political theory so that they encompass the non-mathematical arguments of section 2 but exclude the model-building and theorem-proving exercises of section 3?

In section 2, I argued that Pettit operates with two distinct notions of popular control. According to one, popular influence imposes an equally acceptable direction on policy; according to the other, popular influence imposes a direction on policy that is compatible with universally accepted policy-making norms. This argument was clearly an exercise—whether or not one judges it successful—in political theory, or "normative political theory," as traditionally understood in Anglophone political science and philosophy departments. Drawing distinctions between concepts and comparing their implications are political theory's bread and butter.

In section 3 I presented a model of norm-based popular control that consisted of formal definitions of policy-making norms, citizens' attitudes towards them, and the causal relationship between their attitudes and the direction of government. These definitions were "formal" in the sense that they were cast in terms of sets and functions and introduced special symbols for these abstract objects. But the aim of the model was much the same as the aim of the non-mathematical inquiry of the previous section. Throughout the goal was to explore the implications of government's subjection to accepted norms. In section 2 we considered—without the help of the model—what this subjection implied

about the general acceptability of government policy, whereas in section 3 we considered—with the help of the model—what it implied about the responsiveness of government to citizens' normative attitudes.

Admittedly, the model-building exercise of section 3 does not *look* like political theory as traditionally practiced. But its use of special symbols marks no more than a superficial difference with the intellectual exercise of section 2. One could rewrite section 3 using only the 26 characters of the alphabet. The more economical expressions that the special symbols make possible are mental aids, dispensable in principle but indispensable in practice. Representing claims about popular control and responsiveness with sets and functions, and introducing special symbols for these objects, helped us follow and record their logical implications. What made the model useful was not the presence of numbers—there was nothing quantitative about our topic. Its utility derived instead from the logical complexity of the subject matter. If there is a place for this level of complexity in political theory, then there is a place for formal models and theorems, too.

Appendix

Theorem 1 can be established by proving that it is a corollary of the Gibbard-Satterthwaite (G-S) theorem.[6] Let $\mathbb{P}$ denote the set of complete, transitive binary relations—"preference orderings"—on X. We will refer to generic elements of $\mathbb{P}^n$ with the notation $\succsim$ or $\succsim'$, using $\succsim_i$ to denote the ith component of the n-tuple.

[6] An alternative, independent proof is available upon request.

The G-S theorem states that if $|X| \geq 3$, there is no function $\phi : \mathbb{P}^n \to X$ that is strategy-proof, surjective, and non-dictatorial (as defined below).[7] Theorem 1 is a corollary, because its negation implies the existence of such a function, contrary to the G-S theorem.

To prove this, let us first define a function f that maps from $\mathbb{P}^n$ to $\mathbf{P}^n$. The ith component of $f(\succsim)$, which we denote by $f(\succsim)_i$, will be defined as the collection of sets, $\{Q_i^j\}_{j=1}^{|X|}$, constructed as follows:

$$Q_i^1 := \{x \in X | \forall y \in X, x \succsim_i y\},$$

and for $j = 2, \ldots, |X|$,

$$Q_i^j := \{x \in X | \forall y \in X \setminus Q_i^{j-1},\ x \succsim_i y\}.$$

Since $\succsim_i$ is complete and transitive, $Q_i^1 \neq \emptyset$ and $Q_i^{j-1} \subseteq Q_i^j$ for $j = 2, \ldots, |X|$ and $Q_i^j = X$ for some $j \leq |X|$. (The first set Q_i^1 contains the alternatives tied for first place; the jth set Q_i^j contains the alternatives tied for j-th place plus all the higher-ranked alternatives.)

A function $\phi : \mathbb{P}^n \to X$ is said to be *strategy-proof* if, for any $k \in N$, we have

$$\phi(\succsim) \succsim_k \phi(\succsim')$$

whenever $\succsim$ and $\succsim'$ are identical with the possible exception of their kth com-

[7] For a rigorous but accessible statement, proof, and explanation of the theorem, see Austin-Smith and Banks (2005).

ponent. It is *dictatorial* if there is a $k \in N$ such that, for all $\succsim \in \mathbb{P}^n$, $\phi(\succsim) \succsim_k z$ for all $z \in X$. It is *surjective* if for every $x \in X$, there is a $\succsim \in \mathbb{P}^n$ such that $x = \phi(\succsim)$.

Theorem (Gibbard-Satterthwaite)**.** *Assume* $|X| \geq 3$. *If* $\phi : \mathbb{P}^n \to X$ *is surjective and strategy-proof, then it is dictatorial.*

To see the connection between theorem 1 and the G-S theorem, observe that given functions $g : \mathbf{P}^n \to X$ and $f : \mathbb{P}^n \to \mathbf{P}^n$, the composite function $g \circ f$, defined by $g \circ f(x) := g(f(x))$, maps preference orderings to outcomes in X, so the G-S theorem tells us that if $g \circ f$ is strategy-proof and surjective, then it is dictatorial. Theorem 1 is a corollary of the G-S theorem because it can be shown that if theorem 1 were false, then there would exist a $g : \mathbf{P}^n \to X$ such that with f defined as above, $g \circ f$ is strategy-proof, surjective, and non-dictatorial.

Lemma 1. *If* $g : \mathbf{P}^n \to X$ *satisfies positive responsiveness and popular control but is not dictatorial, then* $g \circ f$ *is strategy-proof and surjective but not dictatorial.*

Proof: assume the hypothesis. Let any $x \in X$ be given. Take any $\succsim \in \mathbb{P}^n$ such that, for all $i \in N$, $x \succ_i z$ for all $z \in X \setminus \{x\}$. Then, $\{x\} \in f(\succsim)_i$ for all $i \in N$. Thus, $g(f(\succsim)) = x$, by popular control. Hence $g \circ f$ is surjective.

To see that $g \circ f$ is not dictatorial, assume the contrary and let $k \in N$ be the dictator. From this assumption we will show that g is dictatorial, contradicting the hypothesis. To see that k is a dictator under g, we choose an arbitrary $\mathcal{P}' \in \mathbf{P}^n$ and any $x \in \cap_{P \in \mathcal{P}'_k} P$ and show that $g(\mathcal{P}') \in \cap_{P \in \mathcal{P}'_k} P$. If $g(\mathcal{P}') = x$, we are done, so let $g(\mathcal{P}') = y \neq x$. Now consider a $\succsim \in \mathbb{P}^n$ such that $x \succ_k z$ and

$z \succ_i x$ for all $z \in X \setminus \{x\}$ and $i \in N \setminus \{k\}$. Since $g \circ f$ is assumed dictatorial, $g(f(\succsim)) = x$. But now, by positive responsiveness of g, there must be some $j \in N$ such that either (i) there is a $Q \in \mathcal{P}'_j$ such that $y \in Q$ and $x \notin Q$, but for all $P \in f(\succsim)_j$, $y \in P$ only if $x \in P$; or (ii) there is a $Q \in f(\succsim)_j$ such that $x \in Q$ and $y \notin Q$, but for all $P \in \mathcal{P}'_j$, $x \in P$ only if $y \in P$. From the specification of $\succsim$ and the definition of f, we know that neither (i) nor (ii) is true of any $j \in N \setminus \{k\}$. From $x \in \cap_{P \in \mathcal{P}'_k} P$, we know (i) is not true of k. So we conclude that (ii) is true of k. But this means $g(\mathcal{P}') = y \in \cap_{P \in \mathcal{P}'_k} P$, because $x \in \cap_{P \in \mathcal{P}'_k} P$. Thus, k is a dictator under g, yielding the desired contradiction and proving that $g \circ f$ is not dictatorial.

To show that $g \circ f$ is strategy-proof, let any two $\succsim, \succsim' \in \mathbb{P}^n$ be given such that $\succsim_i = \succsim'_i$ for all $i \neq k$. We wish to show that $g(f(\succsim)) \succsim_k g(f(\succsim'))$. If $g(f(\succsim)) = g(f(\succsim'))$, we are done, so assume $g(f(\succsim)) = x$ and $g(f(\succsim')) = y$ ($x \neq y$). Then, by positive responsiveness and the fact that $\succsim_i = \succsim'_i$ for all $i \neq k$, it must be that either (i) for all $P \in f(\succsim)_k$, $y \in P$ only if $x \in P$; or (ii) there is a $Q \in f(\succsim)_k$ with $x \in Q$ but $y \notin Q$. In either case, the definition of f implies that $x \succsim_k y$. Hence, $g \circ f$ is strategy-proof.□

Theorem 1 now follows from the lemma and the G-S theorem.

References

Arneson, Richard J. 1999. "Against Rawlsian Equality of Opportunity." *Philosophical Studies* 93(1):77–112.

Arrow, Kenneth. 1951. *Social Choice and Individual Values*. New York: John Wiley & Sons, Inc.

Austin-Smith, David and Jeffrey Banks. 2005. *Positive Political Theory II: Strategy and Structure*. Ann Arbor, MI: University of Michigan Press.

Gibbard, Allan. 1973. "Manipulation of voting schemes: a general result." *Econometrica* 41(4):587–601.

Ingham, Sean. Forthcoming. "Social Choice and Popular Control." *Journal of Theoretical Politics* .

Patty, John and Elizabeth Maggie Penn. 2014. *Social Choice and Legitimacy: The Possibilities of Impossibility*. Cambridge: Cambridge University Press.

Pettit, Philip. 2013. *On the People's Terms: A Republican Theory and Model of Democracy*. Cambridge: Cambridge University Press.

Rawls, John. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Rawls, John. 1999. *A Theory of Justice*. revised edition ed. Cambridge, MA: Harvard University Press.

Satterthwaite, Mark Allen. 1975. "Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions." *Journal of economic theory* 10(2):187–217.

Sen, Amartya. 2014. The Informational Basis of Social Choice. In *The Arrow Impossibility Theorem*, ed. Eric Maskin and Amartya Sen. New York: Columbia University Press.